

RECEIVED
CENTRAL FAX CENTER

MAR 27 2006

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Donna L. Robinson Docket No.: S-100,543
Serial No.: 10/656,358 Examiner: B. J. Forman
Filed : 9/4/2003 Art Unit: 1634
For: : IMPROVED METHODS FOR SEQUENCING GC-RICH AND CCT
REPEAT DNA TEMPLATES

DECLARATION OF DONNA L. ROBINSON UNDER 37 CFR 1.132

I, Donna L. Robinson, hereby state and declare:

1. I am the inventor of the invention described and claimed in the above-referenced patent application (the "subject invention").

2. The subject invention was developed in response to difficulties being experienced by the Los Alamos National Laboratory (LANL) team of researchers working on sequencing the human genome, as part of the Human Genome Project, of which I was a member. After the completion of the "draft phase" of the Human Genome Project, the LANL team became a "Finishing Team" as part of the DOE Joint Genome Institute. LANL's Finishing Team worked primarily on closing "gaps" on Chromosome 16. These gaps were a result of the difficulty of generating and collecting high quality sequence data in various segments of chromosomal DNA because of regions of high G-C content and repeat sequences, which sometimes resulted in secondary structures resistant to sequencing using standard methodologies.

3. While the primary responsibility of the LANL Finishing team was to close gaps presented by these difficult regions on Chromosome 16, the responsibility of a subset of LANL's Finishing Team (the R&D Team) was charged with determining the best (most effective) method of sequencing through GC-rich sequence

S-100,543 Robinson Declaration

samples. The R&D Team tried many different approaches to reading through these difficult regions, including the application of various methods designed to more effectively sequence high G-C content segments, as well as different techniques and tricks in the literature. In addition, the R&D Team tried over a dozen different commercially available kits for sequencing high G-C content DNA. Notably, the R&D Team tried the procedure outlined in the Roche reference cited by the Patent Office in its October 26, 2005 Office Action, but was not successful using it. The R&D Team also attempted to apply commercially-available additives, different sized fluorescent dyes, and dUTP towards the problem. Essentially, none of the methods utilized were effective at providing accurate sequence information across these difficult regions.

4. I was the team leader for the "Production Sequencing Team" that would ultimately employ the determined "best method" in our production sequencing line. I had a lot of experience in sequencing, and a few ideas I wanted to test myself, so I asked my supervisors permission to test them out independently and in parallel to the R&D Team's efforts. After a considerable amount of effort, involving numerous experiments in which all sequencing conditions were pushed to their limits and conventional thinking and methodologies were ignored, I was able to develop a set of conditions that proved surprisingly effective at sequencing through difficult regions characterized by high G-C content and/or the presence of CCT repeats. Overly simplified here, my approach basically involved straining the parameters, exploiting the relationships between components and conditions, and generally pushing the limits in order to provide the best chance for retaining an open configuration in the DNA to be sequenced, for as long as possible, so that the polymerase used in the sequencing reaction could read-through the difficult region before the extreme conditions imposed on the enzyme would render it ineffective. A principal component of my thinking was that I needed to select high Td primers capable of functionally annealing at the much higher temperatures I wanted to use to maintain an open template conformation.

S-100,543 Robinson Declaration

5. Attached as Exhibit A is a true copy of a part of the "Technical Description" section of the Invention Disclosure for the subject invention, which I prepared and submitted to patent counsel for LANL on November 13, 2002. In this section of the Invention Disclosure, I provided additional details concerning the development of the claimed methods.

6. The results obtained with my so-called "GC Buster" method were surprising and unexpected. Despite pushing conditions substantially beyond what the art accepted as viable at the time, the methods worked better than any of the other approaches being attempted by the R&D Team. As a result, the invention became a critical element of closing the gaps in a number of very difficult regions. Attached as Exhibit B is a collection of e-mail communications within the Finishing Team that attest to the successful use of the invention to close gaps in difficult regions.

7. One of the most unexpected aspects of the results obtained with the methods of the invention was the very high quality of the sequence information over exceptionally long read-lengths. The data presented in FIGS. 1-4 of the subject application compared the methods of the invention to modified standard sequencing conditions. Briefly, the method of the invention (utilizing high Td primers in combination with high temperature cycling conditions) was compared to the use of high temperature cycling conditions alone. The method of the invention out-performed the modified standard sequencing protocol in all cases. In one comparison (see Example 2, page 30, and FIGS 3A and 3B), the method of the invention achieved 99% confidence level quality base reads over 411 bases, compared to only 116 bases using the high temperature cycling conditions with standard primers. Moreover, the use of high temperature cycling conditions alone resulted in a complete loss of quality data beyond template residue 330. In contrast, the use of high temperature cycling conditions in

S-100,543 Robinson Declaration

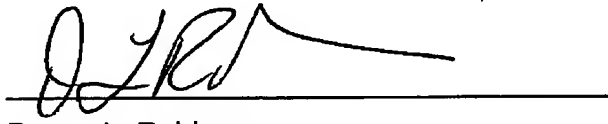
combination with the high Td primers of the invention generated excellent data through about template residue 600.

8. At counsel's request, I have recently reviewed the Roche reference cited by the Patent Office in rejecting the subject application's claims. As noted above, the procedure disclosed in Roche was one of the many failed approaches that the Finishing Team applied to the problem of finishing sequences in regions of high G-C content or containing CCT repeats.

9. The Roche reference describes a protocol that is quite different from the claimed methods. Roche is completely silent on the use of high Td primers, a critical component of the claimed methods, as indicated in paragraph 7, above. In addition, Roche's procedure call for annealing at a temperature of 45-65°C, compared to 65-67°C in the claimed methods. Further, Roche's procedure calls for extension conditions to be run at either 68 or 72°C, for a time period calculated by 45 seconds per kb of DNA to be amplified. In contrast, the claimed method (claim 1) requires extension at a higher temperature range (75-78°C) for a much longer time (3-4 minutes). Thus, the extension conditions are quite different from those disclosed in Roche. In a sequencing reaction over, for example, 600 bases of G-C rich template DNA, extension times of only 45 seconds (or less) would be ineffective at generating the sequencing information desired using the method of the invention. This is presumably a consequence of the strain placed on the polymerase at such high extension temperatures, resulting in a slower enzymatic activity. In designing the extension conditions of the invention, I wanted to provide the enzyme with substantially more time to counteract the strain placed on the enzyme by higher temperatures in order to achieve longer read-lengths. Empirical data showed that about 3-4 minutes was required for successful operation of the method. Overall, I see little similarity between the procedure described in the Roche reference and the methods of the invention. Like the numerous other approaches taken by the Finishing Team, the Roche method failed to solve our G-C rich sequencing problems.

S-100,543 Robinson Declaration

10. All statements made herein of my own knowledge are true and all statements made on information and belief are believed to be true and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.


Donna L. Robinson3-27-06
Date

DECLARATION OF DONNA L. ROBINSON

EXHIBIT A

Appendix A

2. Invention Description And Commercial Potential**"GC-BUSTER SEQUENCING METHOD"****a) General Purpose:**

The invention is a method to generate sequence data in regions of genomic DNA that are heavy in guanine-cytosine (GC-rich (with or without secondary structure)) and CCT repeats, the types of regions that, until now, researchers have not been able to sequence. The method will enable researchers to contribute directly toward completing the sequencing of the human genome, and it will enable the sequencing of the GC-rich (with or without secondary structure) and CCT repeat regions of all organisms (i.e., animals, plants, and fungi). Immediate applications for this method include forensic and clinical-based projects where having a complete set of genetic sequence information is crucial for the analysis and the outcome of such projects. Such projects include research funded by DOE, including those in threat reduction, and pharmaceutical research.

b) Technical Description, Part A:

The ability to generate (to make a fluorescent copy of a template DNA sample) sequence data from GC-rich (with or without secondary structure) and CCT repeat regions of DNA samples has been an almost insurmountable problem faced by scientists working on the Human Genome Project for the past four years. Scientists have found that the genomic information in humans that is represented by these GC-rich (with or without secondary structure) and CCT repeat regions could contain the coding information that is crucial for transcribing genes. The immediate importance of successfully generating and collecting this genomic information can be summarized as follows:

- (1) High GC-rich, without or without secondary structure, and CCT repeat regions are expected to have coding regions within them that the genome community is required to have accurately sequenced.
- (2) The ability for DOE scientists to meet the established finishing goals set and funded by the DOE depends on success in generating and collecting these data.
- (3) Future sequencing projects, such as those supporting our nation's security (threat reduction), and other forensic and clinical-based projects will require having a complete set of genetic sequence information for analysis.
- (4) Scientists believe that the rate of occurrence of these types of GC-rich, with or without secondary structure, and CCT repeat regions may actually occur at a much higher rate in many other organisms than has been seen in humans. Therefore, having a developed method for generating and collecting this data for future projects should prove fundamental to the success of these types of projects.

The method was conceived in an effort to support a project headed by Mark Mundt, PI for the Informatics Group at the Center for Human Genome Studies at LANL (CHGS/LANL). The success of meeting the finishing goals on the project included and required finding a way to collect accurate sequence data in GC-rich, with or without secondary structure, and CCT repeat regions. After forming the idea for the method, I asked Mark Mundt for permission to test my idea in parallel with all the efforts being put forth by the R&D group for the CHGS/LANL to address this same problem. The R&D group tried approximately 20 different methods, either using commercially available products that are marketed as being effective for these GC-rich sequences, or employing methods found in journal articles where different approaches for successfully producing Polymerase Chain Reaction (PCR) products from GC-rich DNA samples were described. Some of these methods, which include commercial methods (additives) and methods described in journals, are described below in Appendix A, "Sampling of some of the Methods used by the R&D Group."

With Mark Mundt's support, I began to test and develop my idea or method of generating sequence data on samples that were known to be GC-rich (with or without secondary structure) in January 2002. The basic premise of my idea was based on understanding and defining (determining the limits) the basic nature of each of the components and conditions in a sequencing reaction, exploiting the relationships among components and conditions, and then pushing their limits to effectively drive the sequencing (the generation of a fluorescent copy) through these difficult regions to collect the sequence data.

DNA, the substance of genes, is composed of four basic building blocks, or bases: guanine (G), adenine (A), thymine (T), and cytosine (C). These four bases are arranged like a chain--in a tandem order--to form a DNA strand. This tandem order of bases constitutes basic genetic information (sequence data). There are two strands of bases that are parallel and complimentary to each other. These strands are bonded (or connected) by hydrogen bonds between the complimentary base pairs, resulting in the formation of the DNA double helix (similar to a twisted ladder). The base guanine always pairs with the base cytosine (GC), and has 3 hydrogen bonds. Adenine always pairs with thymine (AT) and has 2 hydrogen bonds. To collect genetic information (sequence data) on DNA samples, researchers generate a synthetic fluorescent copy from one the strands of DNA that is serving as a template.

Temperature plays an important role in dissociating (separating) the double helix arrangement of a DNA sample to obtain template DNA (a single strand of DNA). When determining dissociation temperatures (Td) to characterize (define) a DNA strand, the higher the GC content of the sequence, the higher will be the Td. This higher Td is a result of the requirement for additional heat (energy) needed to break and dissociate the 3 hydrogen bonds between the GC pairs vs. less heat needed to break the 2 hydrogen bonds between adenine and thymine (AT). Because of the 3 bonds holding each GC pair, the long stretches of GC pairs hold tightly together making it difficult to effectively maintain the dissociated state (separation) of the template DNA throughout the cycle sequencing process used to generate a synthetic fluorescent copy of the template DNA.

Conventional cycle sequencing is completed through cycling through several steps. The temperature is changed to allow different steps in the sequencing reaction to take place (See below, **b) Technical Description, Part B:** for detailed information):

1. At the start of the reaction, the temperature is raised to 92 degrees C for 1 minute to allow the template DNA to be dissociated. This step allows the primers and sequencing enzyme to incorporate synthetic fluorescent labeled bases (G,A,T,C) that generate the copy of the template.
2. The temperature is dropped to 50 degrees C for 5 seconds for primer annealing (primer annealing provides an attachment point for the dGTP BDTv3 enzyme to incorporate the synthetic fluorescently-labeled bases).
3. The temperature is raised up to 60 degrees C for the extension step of the reaction.

In developing my new method, I theorized that the heavy GC-rich regions could possibly reassociate when the temperature was lowered, as is done in Step 2 of the conventional method described above, making it impossible for the dGTP BDTv3 enzyme (purchased from Applied Biosystems) to move down the template DNA strand in these areas to generate a fluorescent copy. The tighter bonds in these regions also are the cause of the formation of secondary structure (where the DNA fold tightly together as in a tighter coiled spring), again inhibiting the generation of a fluorescent copy to the template DNA in these areas. I thought if I can maintain a higher temperature during the annealing and extension steps of the sequencing (therefore altering Steps 2 and 3 of the conventional method described above), this would create a condition where heavy GC-rich regions of the template DNA would more effectively be dissociated, and remain so, to allow the dGTP BDTv3 enzyme to generate the fluorescent copy of this area in the DNA sample.

To support my theory of raising the temperature during cycle sequencing, I designed my own GC-buster primers that would also have higher Td's (therefore enabling me to raise the annealing temperature). To make the cycle sequencing more robust, I tested the effect of increasing the amount of dGTP BDTv3 enzyme in my reactions to increase (1) the amount of enzyme activity in my reactions and (2) the availability of fluorescent bases for each extension step, thus increasing the probability of incorporating a fluorescent base at each cycle. To generate the longest copy of template DNA possible (i.e., to get the most sequence data possible per reaction), I tested and developed an idea I had to address readlength (length of the synthetic copy of the template). I theorized I might be able to force, or drive, the number of incorporated bases at the extension step further by lowering the molar concentration of the primer I made available in the sequencing reaction. For example, if there were fewer primed templates in the reaction, this would focus or force the result of each extension step to add more bases to each primed template rather than adding fewer bases to many primed templates (ie: in one extension cycle, add 2 bases to 5 primed templates vs. only being able to add 1 base to 10 primed templates). For a more detailed description of my method, see **An example of GC Buster Method** below in **b) Technical Description, Part B.**

In conclusion, I exploited the relationships of components and conditions to effectively and efficiently collect the most sequence data (longest read) possible through GC-rich, with or without secondary structure, and CCT repeat regions. My GC-buster sequencing method was effective where the other methods described in Section C had failed. The new method allowed scientists at the CHGS at LANL to collect data and close gaps where no other genome project teams were able to do to this point. This method is being used in the CHGS at LANL to contribute to finishing the sequencing of the human genome and in microbial sequencing projects.

DECLARATION OF DONNA L. ROBINSON

EXHIBIT B

>
>>X-Sender: u119272@harold-mail.lanl.gov
>>Date: Wed, 16 Jun 2004 10:11:44 -0600
>>To: Donna Robinson <drobin@lanl.gov>
>>From: "Cliff S. Han" <han_cliff@lanl.gov>
>>Subject: Re: statement on GC-buster method
>>X-Scanned-By: MIMEDefang 2.35
>>
>>Donna,
>>
>>GC-buster is currently our major method to finish gaps, low quality
>>regions in high GC area. There are now 2 - 3 high GC genome at our
>>hands. More than half of our finishing targets in these genomes
>>have to be done with this methods. I estimate that about 5 percent
>>of our total reaction will be run with dGTP kit.
>>
>>Thanks.
>>
>>cliff
>>
>>
>>
>>>Hi Cliff,
>>>Can you write a brief statement on exactly how you feel the
>>>GC-buster method I have been using to label you GC-rich sample
>>>plates has been helping you finish your projects? To what degree
>>>does this method contribute towards this effort?
>>>
>>>Thanks, Donna

In

Sender: lynne@lanl.gov
Date: Fri, 01 Feb 2002 16:18:24 -0700
From: Lynne Goodwin <dynneg@lanl.gov>
X-Mailer: Mozilla 4.72 [en] (X11; U; Linux 2.2.14-5.0 i686)
X-Accept-Language: en
To: drobln@telomere.lanl.gov
Subject: more comments

Donna,

I was just looking at a clone called 314O13.
There was a bad stretch from 55308-55370
of lots of secondary structure. Your tests of v62_65 and v63_68
gave the best answer. Next was the BDT!

Great!!!!

Lynne

In

Page 1 of 1

Sender: mom@lanl.gov
Date: Sat, 16 Mar 2002 11:18:21 -0700
From: Mark Mundt <mom@lanl.gov>
X-Mailer: Mozilla 3.0 (X11; U; SunOS 5.8 sun4u)
To: claudie@telomere.lanl.gov
CC: saunders@telomere.lanl.gov, rox@lanl.gov, munk@telomere.lanl.gov,
drobin@telomere.lanl.gov, bruce@telomere.lanl.gov
Subject: 2050B12 also closed

2050B12 is now closed, so extra experiments on this region may be tapered. I am doing one more assembly as some stray tb reads cluttered up one end, so I am unsure of the overall quality of the project, but I think it will not be too bad. We'll see.

Thanks, Donna, for closing this high GC place. It will be interesting to see if any DENS reactions will work in this location.

Mark

In

Page 1 of 1

Sender: mom@lanl.gov
Date: Thu, 23 May 2002 09:26:58 -0600
From: Mark Mundt <mom@lanl.gov>
X-Mailer: Mozilla 4.7 [en] (X11; U; SunOS 5.8 sun4u)
X-Accept-Language: en
To: doggett@telomere.lanl.gov, buck@telomere.lanl.gov,
tesmer@telomere.lanl.gov, drobin@telomere.lanl.gov,
saunders@telomere.lanl.gov, rox@lanl.gov, claudie@telomere.lanl.gov,
lynneg@lanl.gov, munk@telomere.lanl.gov
Subject: 167B4 closed after many trials

All:

I believe this rates as the next oldest big problem in our queue after the success we had in getting 1-8F to finally work. 167B4 has been around since Darrell Ricke was here and was one of our first BACs but never closed because of a terrible CCT form repeat we could never get through. Two days ago, the data arrived from a shatter library on a PCR product (we had tried on what we thought was the only spanning subclone shatters before). This appears to have closed the last gap quite well, and I am predicting closure for real on the next assembly. Please plan on redoing DENS and checking for SNP's soon, and congratulations to this big team for all the efforts on this long-standing issue. I hope it is a prediction of our ability with all of our new techniques to be able to finish these tough types of regions.

Thanks.

Mark

In

Page 1 of 1

Sender: mom@lanl.gov
Date: Mon, 03 Jun 2002 00:03:33 -0600
From: Mark Mundt <mom@lanl.gov>
X-Mailer: Mozilla 4.7 [en] (X11; U; SunOS 5.8 sun4u)
X-Accept-Language: en
To: Patti Willis <willis@lanl.gov>, lynneg@lanl.gov, munk@telomere.lanl.gov,
saunders@telomere.lanl.gov, drobin@telomere.lanl.gov,
doggett@telomere.lanl.gov, rds@lanl.gov
Subject: Re: 377 data transfer, 591M7 and 962B4 not loaded

> Lynne G.

These sequences here are big time. Many will be CCT repeats not able to be sequenced by TIGR in the past. I'd like to know a little more about the PCR products or locations targeted with these primers. We may need to shatter products to really get close to the targets with all the repeats near these tough spots. Please also put any of these through Donna's high temp treatment if possible.

It does appear you have closed the gaps in CTA-67A1 and CTA-363E6, both good stretches of CCT, already.
Congratulations!
CTA-279B10 is also possible but a little more tricky to check the results on, so we may need more quality.

For Liz' info, many of these new TIGR projects are located in /biodata2.

Thanks.

Mark

> The following has been sent to yac:

>

> 052902.plt1.bdt.lag.464

> 052902.plt2.bGTP.lag.463

>

> Thanks, Patti

>

> Sorry Mark - didn't sent email to you yesterday Thursday 5/30

>

> Patti